REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE July 1995	3. REPORT TYPE AND DATES COVERED Final Report July 1995

4. TITLE AND SUBTITLE

Database System Studies in Fine Grain Optoelectronic Computing

5. FUNDING NUMBERS

N00014-93-1-0414

6. AUTHOR(S)

Professor Sadik C. Esener

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of California, San Diego
Dept. of Electrical and Computer Engineering
9500 Gilman Drive
La Jolla, CA 92093-0407

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street
Arlington VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release;
Distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

We have studied the relational database architecture for a bi-orthogonally accessed 2-photon 3D memory system. The performances of various database operations, such as projection and selection, have been evaluated based on the Wisconsin benchmark. Effects of data packing and record fill factor have been investigated. The packing strategy was found to have large effect on performance. The effect would be reduced for larger relation sizes. No particular data packing strategy nor word size was found to be clearly advantageous for all operation.
We also evaluated various optoelectronic technologies for parallel optoelectronic computing. The areal data throughput and the energy per transmitted data bit were used in the comparison of the technologies. It was found that, based on current technology parameters, PLZT modulator and VCSEL technologies are well suited for applications in which a large fan-out is required but the total number of transmitters is relatively small. MQW modulator technology is a good candidate for applications in which many transmitters with a limited fan-out are needed.

DTIC QUALITY INSPECTED 3

15. NUMBER OF PAGES
19 (Figs. included)

14. SUBJECT TERMS

VCSEL tech/PLZT mod/MQW mod

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT

Final Report

for

Database System Studies in Fine Grain Optoelectronic Computing

Grantee

The Regents of the University of California
University of California, San Diego
La Jolla, CA 92093-0407

19960822 239

Principal Investigator: Sadik C. Esener
                                (619) 534-2732

Program Manager: William Miceli

Program Monitor: John Starcher

Table of Contents

List of Figures and Tables

filling field parallel pages first. For field-field access packing record parallel pages first yields the worse performance. These graph neglects the effect of fill factor and seek time.

Table 1   Variables used for performance calculations.

Table 2   Effect of w on P_b, r' and r.

Attachment

A reprint from Applied Optics, paper titled "Digital free-space optical interconnections: a comparison of transmitter technologies".

FINAL REPORT

The main objective of this program was to investigate the design and the optoelectronic implementation of a high performance optical memory-processor interface for database applications. For very large database machines, in general, memory bandwidth is a bottleneck. Typically, the access time of a conventional secondary storage devices such as magnetic disks is at millisecond scale. The use of parallel access optical storage systems, such as parallel read-out optical disks and 2-photon 3D memories, have the potential of achieving enormous throughput (> 100 Gbits/sec) and capacity (~ 1 Tbits). In this research, we have studied the relational database architecture based on a bi-orthogonally accessed 2-photon 3D memory. Specifically, the database operation considered is the data filtering. Various optoelectronic technologies have been evaluated for interfacing parallel optic and electronic systems. The system performances have been measured in term of the areal data throughput and the energy required per transmitted data bit.

I. Optoelectronic Database Filter

Database data filters are computers used in database machines to improve the machine performance by eliminating data that is not relevant for a given query when it is retrieved from the secondary storage. By doing this the amount of data transferred to the main machine is reduced, so are the computations necessary for the queries. The optoelectronic data filter examined in this work utilizes a bi-orthogonally accessed 2-photon 3D memory. The data organization scheme is particular to this approach.

1.1 Bi-orthogonally accessed 2-photon 3D memory

2-photon 3D memory devices are made from an organic material, SP-doped PMMA, in which molecules are excited to a high energy state by absorbing two photons. The material is in a shape of either cubic or rectangular. The bits are written throughout the volume by intersecting two laser beams at any bit location at one time. These 3D memory devices also have the feature of being accessed in orthogonal directions, allowing planes of bits which are perpendicular to each other to be retrieved. It is termed bi-orthogonal access.

The concept of bi-orthogonally accessed 2-photon 3D memory for database operation is explained with reference to Figure 1. In one of the orthogonal directions, termed record parallel, the data stored correspond to records of a database. An array of complete records can be retrieved in one memory read-out. Orthogonal to the record parallel direction, the data correspond to a particular field or set of fields of a database, termed field parallel. Figure 1(a) shows the database structure and (b) the bi-orthogonally accessed memory cube.

Data retrieval process is much more efficient in the bi-orthogonal approach, because for a given query data are better isolated. For example in a projection operation, only the field of fields desired need to be retrieved if field parallel access is used, and in certain selection operations scanning can also be performed efficiently with this retrieval mode. If a record is determined to satisfy the selection after a scan , it can be retrieved in one page read with record parallel access. To read out the same record using field parallel access would require many more page reads.

4

1.2   2-photon 3D memory data organization

The data organization scheme in 2-photon 3D memory devices is shown in Figure 2. The memory is divided into $\hat{B}$ super-blocks. A super-block can be viewed as a sequence of pages that can be accessed randomly in either of two orthogonal directions in time T_p. Each super-block is a cube of bits with M bits on a side; thus pages read from these super blocks contain M^2 bits. The total memory capacity is therefore $M^3 \hat{B}$ bits. Records are placed in the memory such that they are contained in one record parallel page, and so each page accessed in the field parallel direction contains w bits of a record. This is shown in Figure 3. With this scheme, a complete record of size r_t can be accessed in one memory read utilizing record parallel access or in roughly r/w reads using field parallel access. The set of field parallel pages containing a complete record is referred to as a *block* and the number of field parallel pages in a block is denoted as P_b. The number of complete blocks in a super-blocks is denoted as b.

It is assumed that data from multiple fields can be contained on the same field parallel page as would occur, for example, when w is large. The parameter r' represents adjusted record size and takes into account the capacity wasted if field parallel pages cannot be not completely filled. This would occur, for instance, if r_t is not a multiple of w. The parameter w affects the fill-factor of the memory in a more significant way if one wants to ensure that a record can always be retrieved with one memory read. This problem, termed record fill factor, is explained with reference to figure 4. Rarely will blocks exactly fill super-blocks. As a result, some field parallel pages will not be used. On average there will be $P_b/2$ such pages per super-block. Since $P_b = r'/w$ this capacity penalty increases when w is small. In these studies the parameter r represents record size taking into account the above two fill factor contributors.

The packing of data in the memory also affects performance. A relation of R records may not completely fill all the super-blocks in which it is contained. Consequently the first and last super-blocks will most likely contain other relations. In this situation, records residing in partially filled super-blocks can placed such that they fill field or record parallel pages first. In general if field parallel pages are filled first, the time to perform operations using record parallel access will approximately increase by a factor of $1/B$, where B is the minimum number of super-blocks required for the operation. This is because on average two additional half super-blocks of data will have to be read that do not contain the relation of interest. With this packing scheme, the time required for operations performed using field parallel access is usually not affected. If, on the other hand, record parallel pages are filled first, the record parallel access time is not affected, but the field parallel access time is increased roughly by a factor of $1/B$. The effect of adverse packing decreases for larger relations, but can be significant for smaller relations.

1.3   Performance Study

In this study the Wisconsin benchmark was selected to evaluate the potential performance of a 3-D two-photon memory based relational database data filter. With this benchmark, the performance of the selection operation is examined for different selectivities. The selectivity of an operation refers to the number of records that satisfy a selection query. A relative selectivity of 10% means that 10% of the records in the relation satisfy the query. An absolute selectivity of 100 records means that 100 records satisfy the query regardless of the relation's size. The Wisconsin benchmark also requires that the performance of selection operations be measured

5

using different types of indexing as well as no indexing. A relation with an indexed field is organized in some way (B-tree, hashing...etc.) according to the value of a particular field or fields; an index key is assigned to a record based on this value. With clustered indexing, the index key determines the physical location of the data.

In the sections that follow, the performance of various selection and projection operations is examined for the 3-D memory based system for the different accessing techniques. To illustrate trends the selectivities are varied continuously and include the discrete selectivities required by the benchmark. Tables 1 and 2 list the parameters and values that were assumed in the performance study. "$\langle \ \rangle$" is used to denote average value; "$\lfloor \ \rfloor$"denotes the integer less than or equal to the operand in brackets; "$\lceil \ \rceil$" denotes the integer greater than or equal to the operand in brackets. $\langle T_{seek} \rangle$ is the average time that it takes to seek to the starting page of an operation. This page is assumed to be unique. The size of the relation was chosen to facilitate comparison with other systems. A single two-photon memory would be able to contain a larger relation.

Projection

Figure 5 shows the time to perform projection neglecting the effect of record fill factor, packing and seek time. This is termed *ideal* projection. Ideal selection assumes the same omissions. The selectivity of the projection operation is augmented in byte increments even though this would physically correspond to reading out fractions of fields. The field parallel mode in general shows superior performance for this operation; unlike the record parallel mode, the entire relation does not have to be read out in the operation, only the field or fields desired. In Figure 6 the effect of seek time and record fill factor are included. The time required to perform projection using field parallel access equals the ideal time only when the size of the field(s) desired is a multiple of w, otherwise additional pages need to be read out. This effect can be seen for $w = 16$ bytes. The time required for this operation using record parallel access is directly related to the record fill factor of the memory. When $w = 1$ byte, roughly 23% more time is required than for the ideal case because r/r_i is equal to 1.23. As w increases the time for this operation using record parallel access decreases since the effect of the record fill factor is reduced. Figure 7 shows how the ideal times for projection are affected with the two different types of packing. As can be seen packing can have a larger effect on performance than record fill factor. With record parallel packing, field access is no longer preferable when a very large portion of a record is desired.

Selection with clustered indexing

For selection with clustered indexing the location of the records is known a priori, and the records satisfying the selection criterion are assumed to be adjacent. The time to perform ideal selection with clustered indexing is plotted vs. selectivity in Figure 8. For this selection operation, a certain minimum number of pages or blocks is required. However, the set of selected records may not completely fill the pages or blocks in which it is contained. This is because the set of records will rarely be an integer number of pages/blocks in size, and also because the set is likely to start in the middle of a page/block. The time required to perform this type of selection with low selectivity is roughly equal to the time to read a single page or block. Field parallel access with $w = 1$ byte, for example, shows particularly poor performance because

6

P_b is so large. For higher selectivity operations, the difference between the selection times reflects the overhead incurred when records do not completely fill pages or blocks. This overhead is once again higher with larger sized blocks. The ideal time to perform this operation with record parallel is always less than for field parallel access, because pages are always smaller than blocks.

When the effect of record fill factor is included, record parallel access is not preferable to field parallel access for all selectivities and all values of w. The time to perform selection with clustered indexing taking into account record fill factor and seek time is plotted in Figure 9. When $w = 1$ byte, the record parallel approach is not preferable to field parallel access for some values of w, because the amount of data that has to be read out is increased due to the record fill factor problem. It should be noted that the two selection times assume a slightly different data ordering: with record parallel access, consecutive records are on the same record parallel page; with field parallel access, consecutive records are on the same field parallel page.

Figure 10 shows the effect of packing on the time required for selection neglecting the record fill factor and the seek time. It shows that the time required to perform selection with clustered indexing is increased when record parallel access is used when the packing strategy involves filling field parallel pages first. Similarly, if field parallel access is used it requires longer time when record parallel pages are filled first.

Selection with no indexing

For selection with no indexing the locations of the records satisfying the selection query are not known a priori. It is assumed that these records are uniformly distributed throughout the memory. Performing selection with no indexing using record parallel access, as with record parallel projection, requires that the entire relation be read out so that each record can be examined to determine which records satisfy the selection query.

Selection with no indexing can be performed using only field parallel access. This is referred to as *field-field* access. Alternatively it can be performed by combining field and record parallel access in what is termed *field-record* access. These two approaches can be broken down into two parts: a search part and a readout part. The search part of the operation is performed using field parallel access. The operand containing field(s) are retrieved using field parallel access. If a field or set of field is determined to satisfy a particular selection criteria, its corresponding record would be retrieved using record parallel access for field-record access or it would be retrieved using field parallel.

In Figure 11 the time to perform ideal selection with no indexing is plotted vs. selectivity for the three access modes. The search time for the field-record plot is assumed to be 310 sec. The time to perform this operation saturates for the two field parallel based accessing approaches. With field-record access this occurs when it is likely that every record parallel page will have a selected record on it and will need to be read out: when the fraction of records satisfying the selection criterion is roughly equal to *(1/(the number of records on a page))*. For field-field access the saturation happens when it is probable that every block needs to be read out: when the fraction of records satisfying the selection criterion is approximately equal to *(1/(the number of records in a block))* -- the larger the block, the quicker the saturation.

Selection times for the record-parallel access mode and the average selection times for the two forms of field-parallel access are plotted in Figure 12 vs. selectivity. Plots for selection implemented with field-field access for small w would be even more vertical than the $w = 16$

7

bytes trace and were not included. These selection operations would also saturate at the ideal record parallel line. For low-selectivity operations field-record access yields the best performance. For selectivities greater than 0.4%, however, field-field access would be advantageous. In this range all blocks and roughly all pages will have at least one selected record on them and will need to be read out. With field-record access, the field containing the operand will have o be readout twice: once while searching and a second time during readout. For $w = 1$ byte, the overhead due to record fill factor additionally hurts the performance with field-record access.

Packing also affects the performance of selection with no indexing. If record parallel pages are filled first, the time to complete the selection operation with pure record parallel access is unaffected as is the time to readout selected records with field-record access. The search time for field-field and field-record access is increased once again by approximately a factor of $1/B$, and the time for readout with field-field access is also increased. The equation for this is given below. It assumes that the first and last super-block are half full. If the packing strategy is to fill field parallel pages first, the time for searching is unchanged as is the time to perform the selection operation with field-field access. The time for record parallel access, however, would roughly increase by a factor of $1/B$, and the time for readout for the field-record approach would be increased as well. The equation describing this increase is given below. Once again it is assumed that the first and last super-block are half full.

The time to perform selection with no indexing is plotted in Figure 13 for the three different accessing modes with and without the effects of worst case packing, neglecting record fill factor. With record parallel and field record access, packing field parallel pages first yields the worst performance unless the search time exceeds the readout time for record parallel access. For field-field access packing record parallel pages first yields the worst performance.

Summary

In this study, no particular packing strategy or word size was found to be clearly advantageous for all operations. For the relation size and data organization strategies considered here, the effect of the packing was found to have a larger effect on performance than the effect of record fill factor. The effect of packing would be reduced for larger relation sizes. However, for smaller relations, it would be increased. In a system, the size of w and the packing strategy would have to be chosen by anticipating frequent operations.

II. Free-Space Optical Interconnects

To interface 2-photon 3D memory devices with electronic processors, we have evaluated various optoelectronic technologies based on free-space optical interconnects (FSOIs). The results of this evaluation have been published in *Applied Optics* (A reprint is attached), and is summarized in this section.

To be able to compare to an all-electronic system, we have defined an optoelectronic interface as shown in Figure 1(rp) (rp indicates the figure in the attached reprint). It begins at the transmitter driver inputs and ends at the detector amplifier outputs. The transmitters can be either light modulators or light emitters. When light modulators are used, an external laser source (P_o) is required. In digital free-space optical interconnection systems, each detector receives the signal from only one transmitter (fan-in = 1). Therefore, for a system with N transmitters, each

with a fan-out of F, the total number of interconnection channels is the product of N and F, refereed to as the connectivity. The connectivity can be expressed as

$$N \times F = \frac{\eta_{os}P_{e,o}}{2\overline{P_d}(BR, BER)}\eta_T(P_e, BR);$$

it is a function of the transmitter power efficiency (η_T), the minimum detectable optical power ($\overline{P_d}$) at the receiver input, the optical link efficiency (η_{os}), and the transmitter driver electrical power (P_e). When light modulators are used, the connectivity is also a function of the input optical power (P_o) to the modulators.

Two important parameters used to evaluate a parallel data transmission system are areal data throughput and energy required for a transmitted data bit. The areal data throughput is the product of the connectivity and the data rate, divided by the required hardware area. The energy per transmitted bit is obtained by dividing the required power (optical or electrical) by the total data throughput. Three transmitter technologies are examined based on the system parameters for various application architectures.

2.1 Transmitter technologies

Three transmitter technologies considered were PLZT modulator, MQW modulator, and VCSEL technologies. Each transmitter technology was evaluated based on its power efficiency. Figures 3(rp), 5(rp), and 6(rp) plot the transmitter power efficiency of the three technologies as a function of the input power, respectively. In the case of PLZT modulators, due to the large device capacitance, the power efficiency is limited by the electrical driving power. For the MQW modulators, on the other hand, the maximum modulated power output depends on the optical saturation intensity. Therefore, the transmitter power efficiency is a function of the input optical power. In the VCSEL case, the power efficiency is related to the electrical-to-optical power conversion efficiency.

2.2 Transmitter Fan-out

By using a common high-impedance optical receiver circuit shown in Figure 10(rp), with the minimum detectable power plotted in 11(rp), the calculated transmitter fan-outs are shown in Figures 12(rp) and 13(a),(b)(rp). In VCSELs, the input power is purely electrical. Once the threshold is reached, the VCSEL's fan-out increases linearly with increasing input electrical power. The maximum fan-out is constrained by the maximum power output of the VCSEL (10 mW assumed in this case). For PLZT and MQW modulators, the fan-out depends on both electrical and optical powers. It is shown in Figure 13(rp) that, with both modulator technologies, there is an optimal ratio of optical-to-electrical power inputs to achieve a maximum power efficiency. In PLZT modulators, this ratio is 2:1 and is independent of the operating data rate. The optical-to-electrical power ratio is approximately 60% for MQW modulators; it is almost a constant up to 2 Gbit/s of data rate.

2.3 FSOI System performances

For a given technology and application specifications, a FSOI system was evaluated by its areal data throughput and the energy required for a transmitted data bit. The areal data throughput is the product of the connectivity and the bit rate, divided by the required hardware area. The maximum connectivity depends on the total available system power and the operating

data rate. The hardware area is determined by the electrical power requirement of the transmitter driver circuits and the maximum power dissipation density. For example, with a fixed connectivity of 16K (128x128), Figure 16(rp) shows the electrical power requirement versus bit rate for several architecture/technology combinations; a total optical power of 10W is assumed for the modulator technologies. The total required hardware area is indicated on the right hand axis by assuming an electrical power dissipation density of $10W/cm^2$. The architectures considered are point-to-point (N=16K, F=1), hypercube (N=1540, $F=Log_2(N) \approx 11$), and crossbar (N=F=128).

Along any vertical line in Figure 16(rp), the data throughput is a constant, the required area increases as the electrical power requirement increases, as indicated on the right axis. For a maximum areal data throughput, therefore, PLZT technology is superior at a data rate below 100 Mbit/s, MQW technology is a good candidate for most applications up to a data rate of 1 Gbit/s, and VCSEL technology is the best choice at a date rate beyond 1 Gibt/s. The limitation on the operation data rate is set by the maximum optical power at the output of the transmitters. With the specified receiver technology, more than 10W of optical power is needed for a modulator system operating beyond 1 Gbit/s, and a 10 mW output power per VCSEL is a necessity. As the array size increases, the power requirement will increase as well. With the same input power, it would require higher receiver sensitivity.

Summary

This study shows that PLZT and VCSEL technologies are well suited for application in which a large fan-out per transmitter is required but the total number of transmitters is relatively small. MQW modulator technology, on the other hand, is good for applications in which many transmitters with a limited fan-out are needed. The limiting factor for the MQW modulator technology in large fan-out applications is its intensity saturation. Whereas the limitation of VCSEL technology for many transmitter applications is its threshold power. In any application, the receiver sensitivity plays very important role in determining the system performance.

student ID	Name	GPA	Major	Address
07839120	B. Jones	4.0	economics	9100 Regents Rd, SD CA 92037
07839121	T. Slash	3.7	engineering	221 Donex Ave, SD CA 92117
07839122	R. Kelsey	2.6	philosophy	3007 Ivy St, SD CA 92410
07839123	K. Smith	2.8	economics	102 Hall St, SD CA 92109
07839124	S. Jensen	3.2	biology	1765 Filbert St, SD CA 92116

field parallel access

(a)

record parallel access

field parallel access

data from single field

record parallel access

record
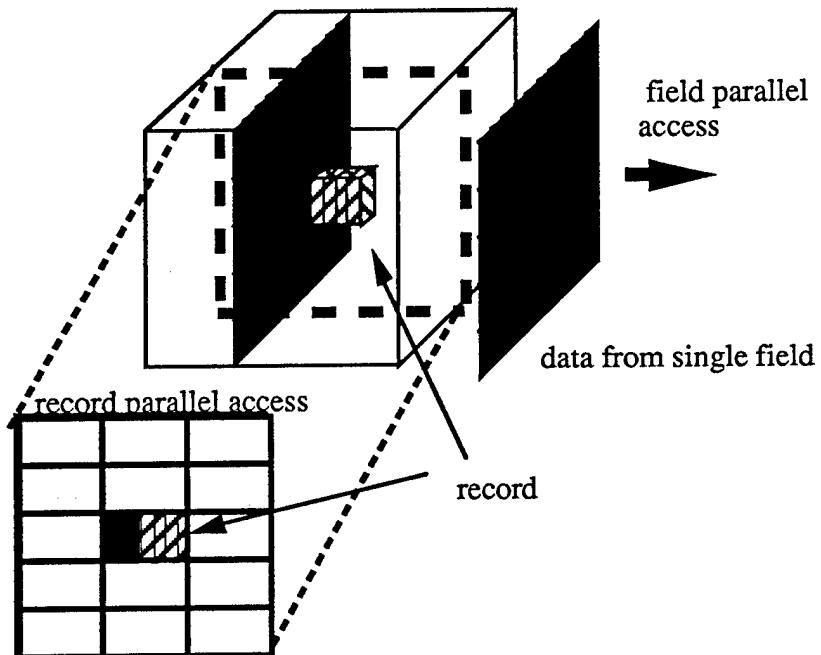
(b)

Figure 1 Data is organized in a bi-orthogonally accessed 3-D two-photon memory such that a page of complete records is be obtained with one memory read using *record parallel* access. Pages containing data from the same field or set of fields may also be retrieved. This is termed *field parallel* access.
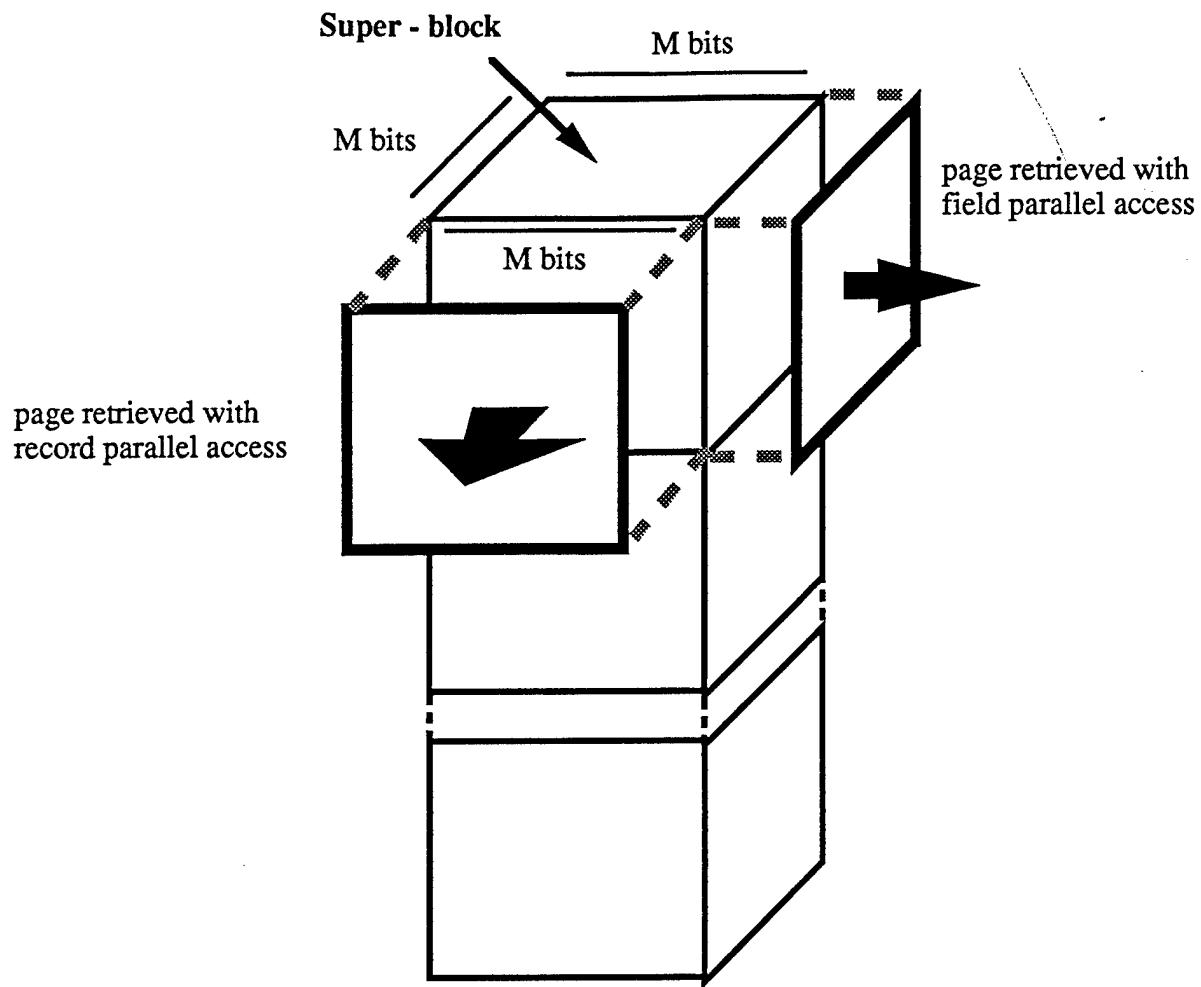
Figure 2 The memory is divided into $\hat{B}$ *super-blocks*. A super-block can be viewed as a sequence of pages that can be accessed randomly in either of two orthogonal direction from any super-block in time T_p. Each super-block is a cube of bits with M bits on a side; thus pages read from these super blocks contain M^2 bits. The total memory capacity is therefore $M^3 \hat{B}$ bits.

Single Super-block:
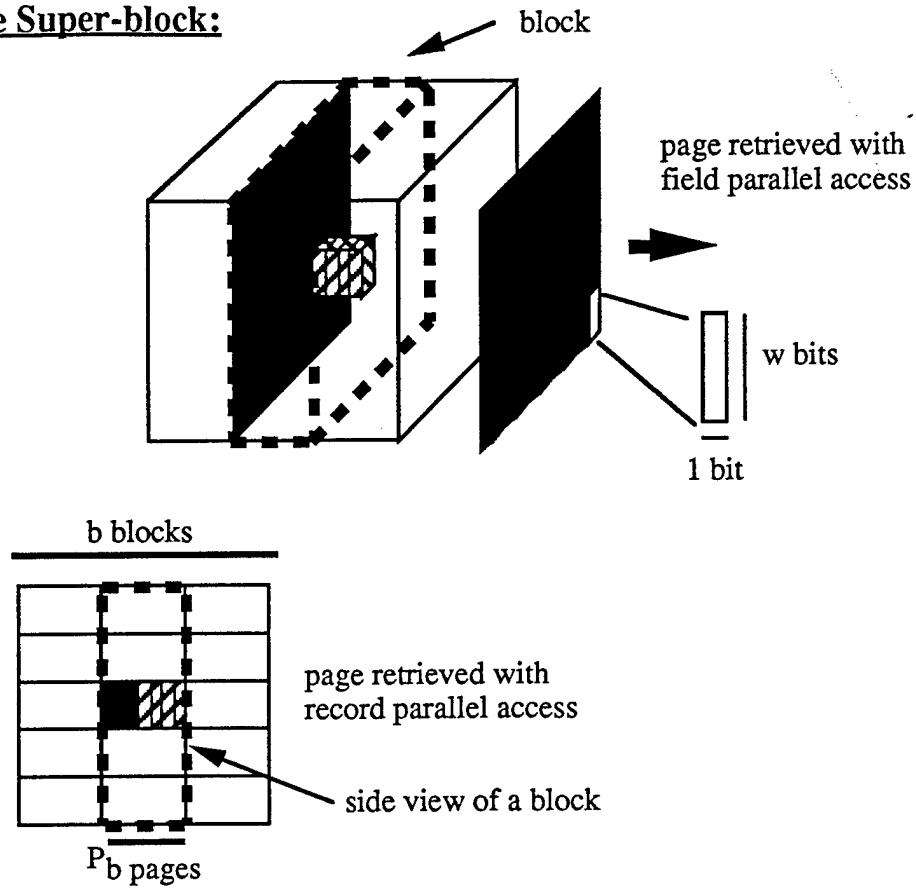
Figure 3   Records are placed in the memory such that they are contained in one record parallel page, and so each page accessed in the field parallel direction contains w bits of a record. A complete record can be accessed in one memory read utilizing record parallel access or in P_b page reads using record parallel access. The set of P_b field parallel pages containing a complete record is referred to as a *block* . Each super-block has b blocks.

Single Super-block:
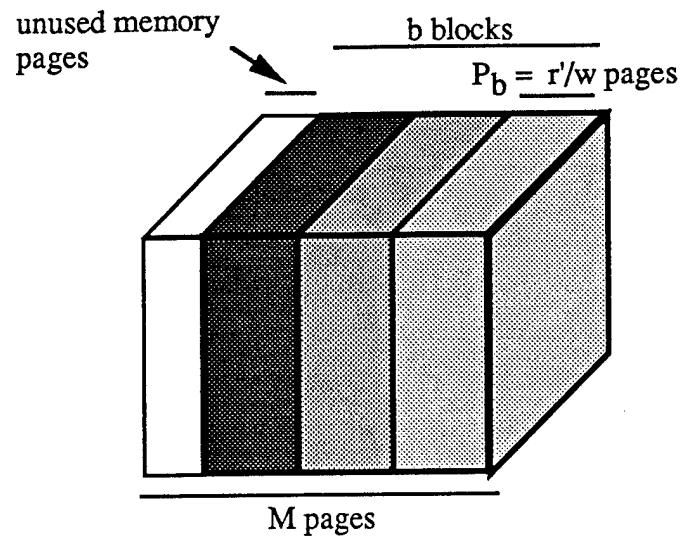
Figure 4 In a system where records can always be accessed in one record parallel page read record fill-factor can become a problem. The length of a super-block, M, is usually not a multiple of P_b, as a result some planes in each super-block will not be used.
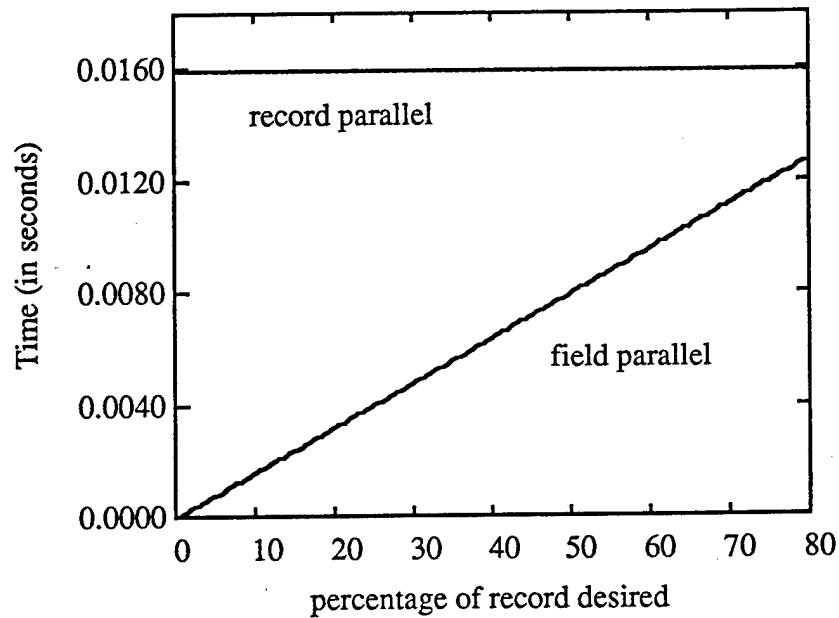
Figure 5 Time required for an ideal projection operation
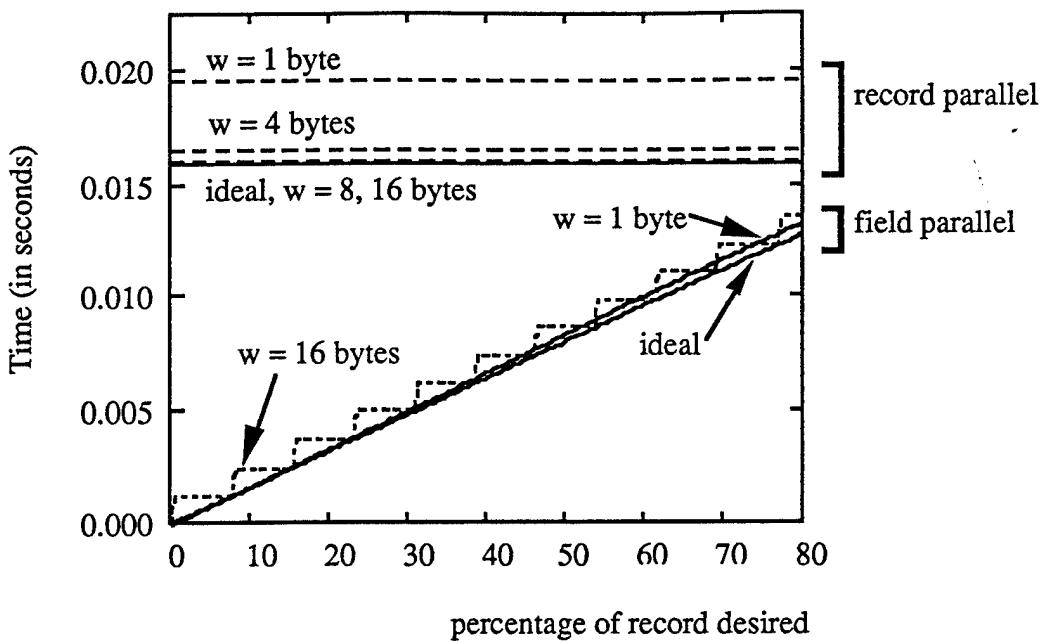
Figure 6 Time required to retrieve the data from the 3-D two-photon memory for a projection operation using field and record parallel access modes. Less time is required for this operation using field parallel access since only the field or set of fields desired need to be retrieved. With record parallel access the entire relation needs to be read out.
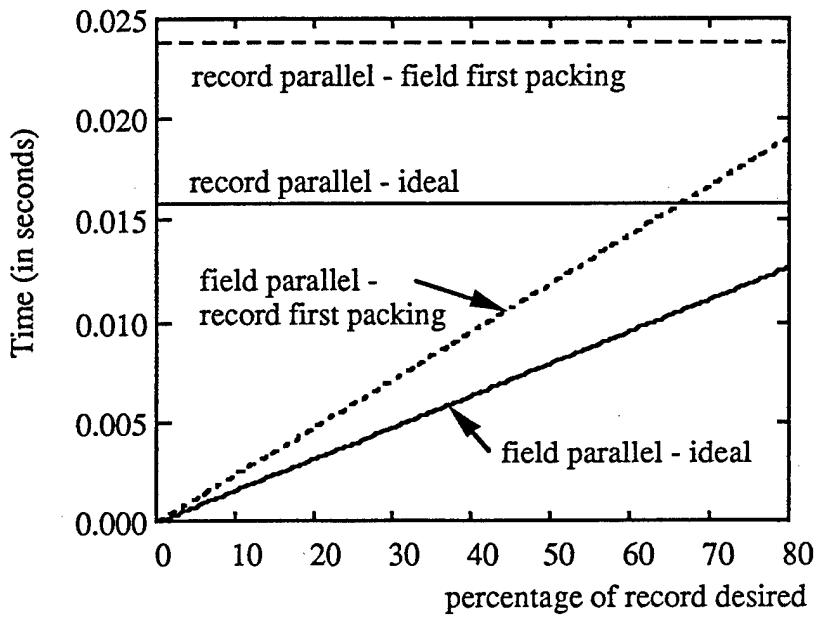
Figure 7 The number of pages that need to be retrieved for a projection operation is increased when record parallel access is used if the packing strategy involves filling field parallel pages first. If field parallel access is used the number of pages is increased if record parallel pages are filled first. This graph neglects the effect of fill factor.
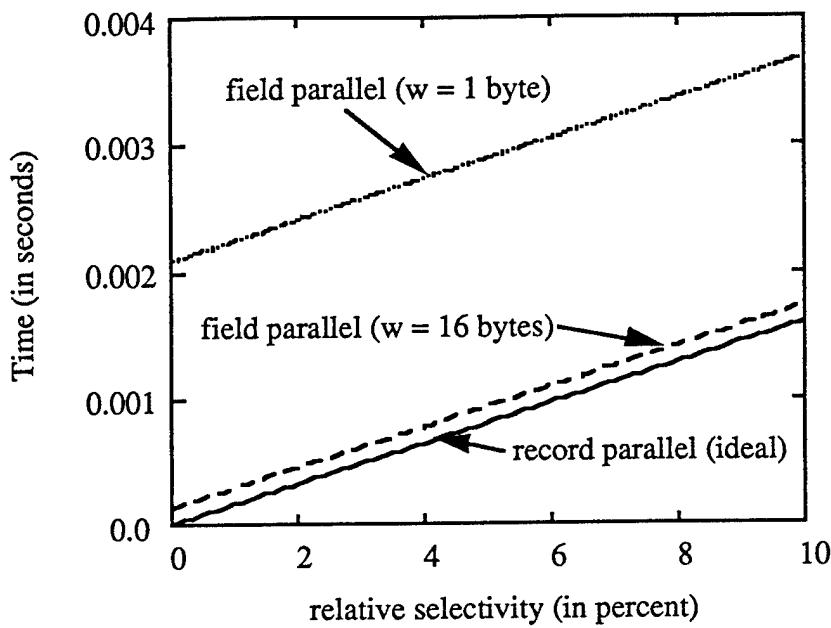
Figure 8 Time to perform ideal selection with clustered indexing
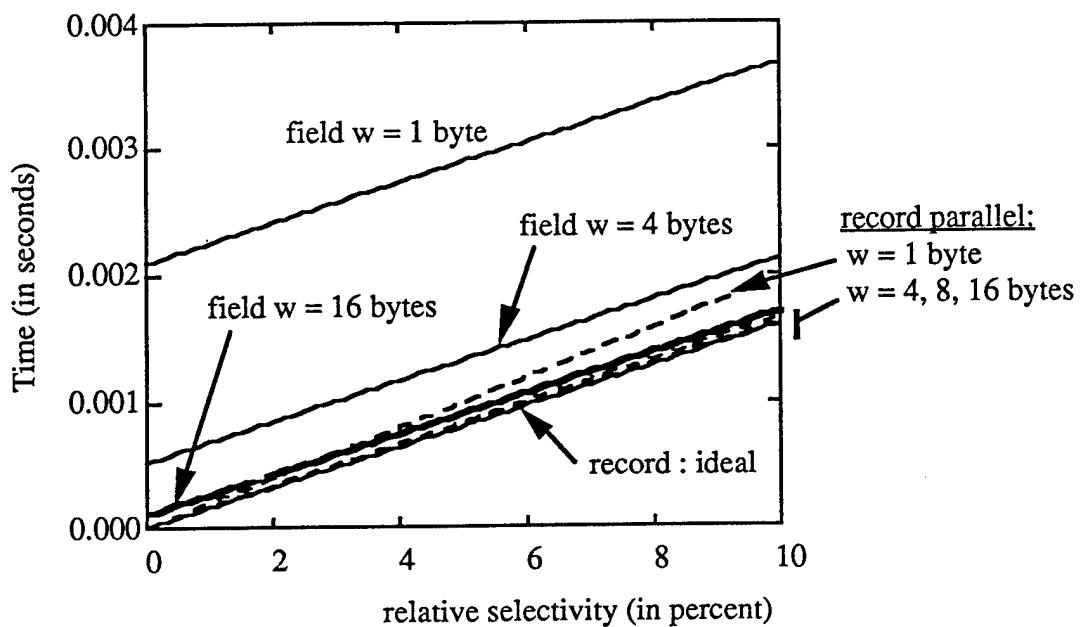
Figure 9 Time to perform selection with clustered indexing neglecting the effect of packing.
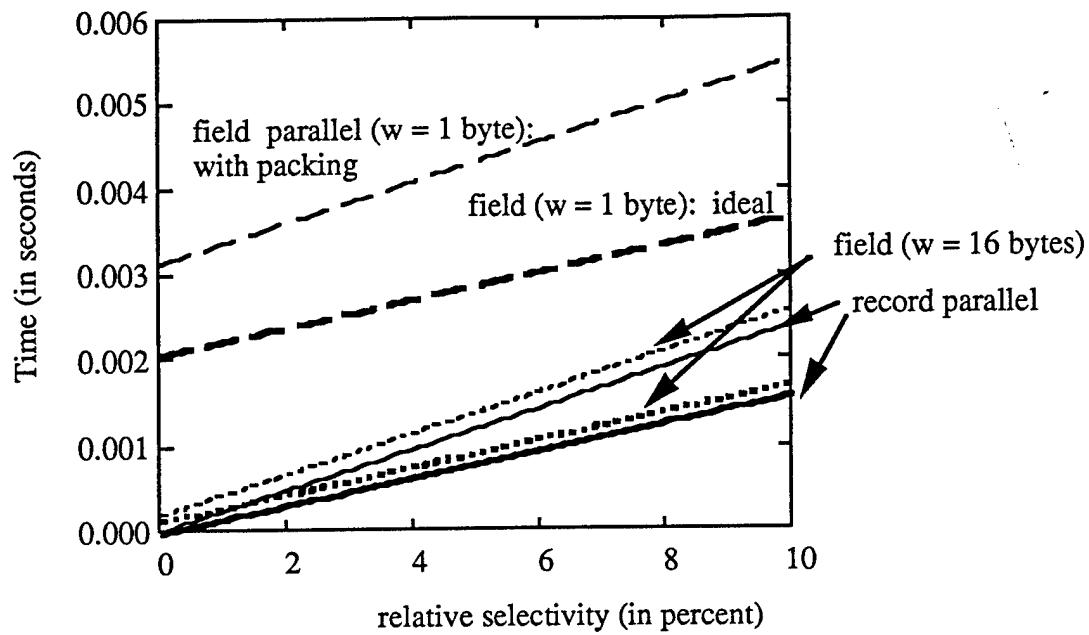
Figure 10 The time required to perform selection with clustered indexing is increased when record parallel access is used if the packing strategy involves filling field parallel pages first. If field parallel access is used the time is increased if record parallel pages are filled first. This graph neglects the effect of fill factor.
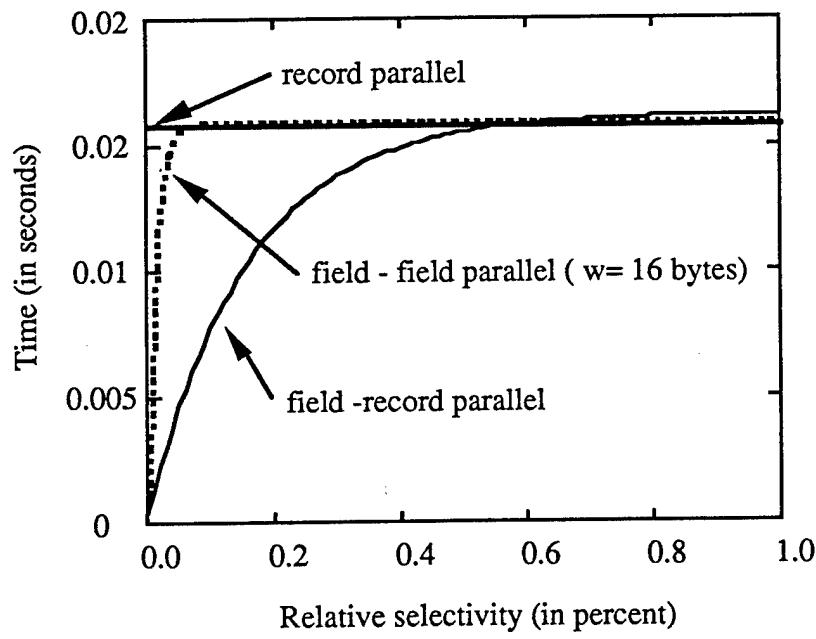

Figure 11 Time to perform ideal selection with no indexing
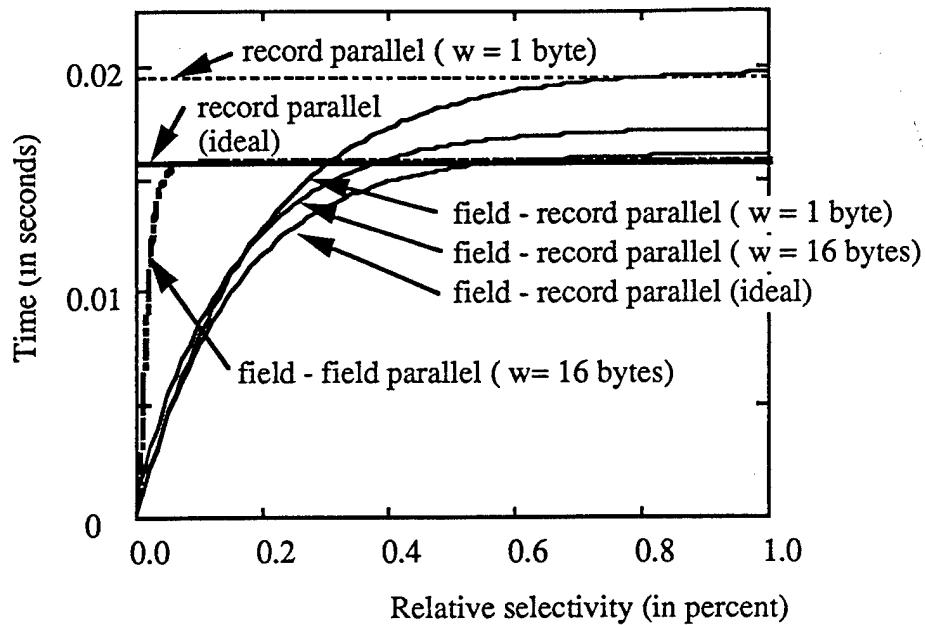
Figure 12 Time to perform selection with no indexing neglecting the effect of packing.
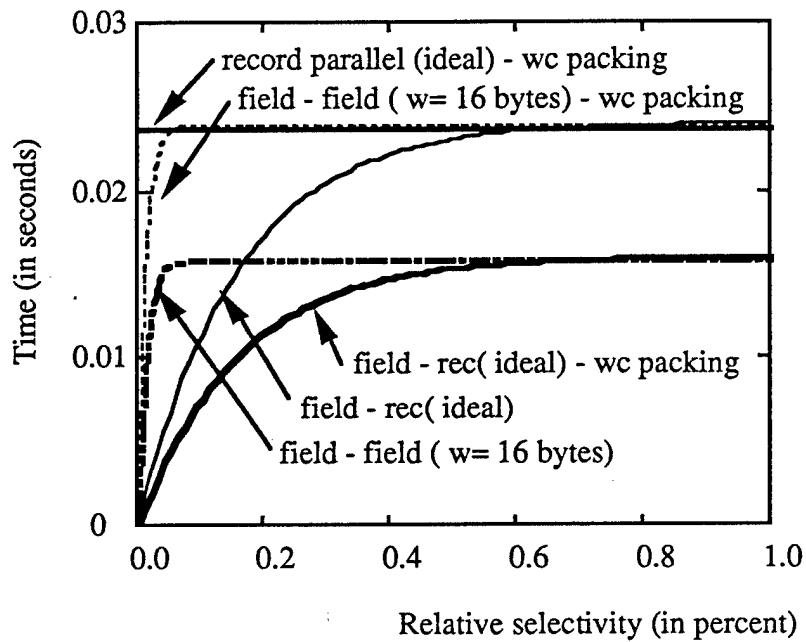
Figure 13 The time required to perform selection with no indexing is affected by packing. With record parallel and field -record access the worst case (wc) packing strategy involves filling field parallel pages first. For field-field access packing record parallel pages first yields the worst performance. This graph neglects the effect of fill factor and seek time.